

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-172680
(43)Date of publication of application : 23.06.2000

(51)Int.CL G06F 17/21
G06F 12/00

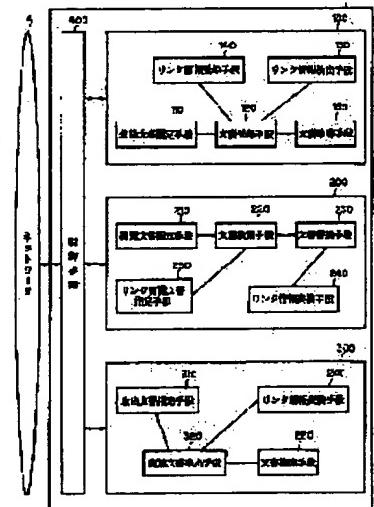
(21)Application number : 10-348299 (71)Applicant : RICOH CO LTD
(22)Date of filing : 08.12.1998 (72)Inventor : IMASATO SHO

(54) SYSTEM AND METHOD FOR REGISTERING DOCUMENT, RECORDING MEDIUM FOR EXECUTING THE METHOD, SYSTEM AND METHOD FOR BROWSING DOCUMENT, RECORDING MEDIUM FOR EXECUTING THE METHOD AND DOCUMENT EXTRACTING SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a system capable of registering, browsing and extracting a hyperdocument capable of tracing a link as it is even after document information to be registered to a document data base is registered to the data base without being changed.

SOLUTION: This document registration system is provided with a registration document designating means 110 for designating the document to be registered, link information detecting means 130 for detecting link information by analyzing the document designated by the registration document designating means 110, document registering means 120 for registering the document, which is made into an object by the link information detecting means 130, to the document data base and link information registering means 140 for managing the link information detected by the link information detecting means 130 and the document registered by the document registering means 120, while relating them.



LEGAL STATUS

[Date of request for examination] 24.02.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C) 1998,2003 Japan Patent Office

BEST AVAILABLE COPY

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号
特開2000-172680
(P2000-172680A)

(43)公開日 平成12年6月23日 (2000.6.23)

(51) Int.Cl. ⁷	識別記号	F I	マーク(参考)
G 06 F 17/21		G 06 F 15/20	5 7 0 D 5 B 0 0 9
12/00	5 4 6	12/00	5 4 6 B 5 B 0 8 2

審査請求 未請求 請求項の数12 O L (全 16 頁)

(21)出願番号 特願平10-348299
(22)出願日 平成10年12月8日 (1998.12.8)

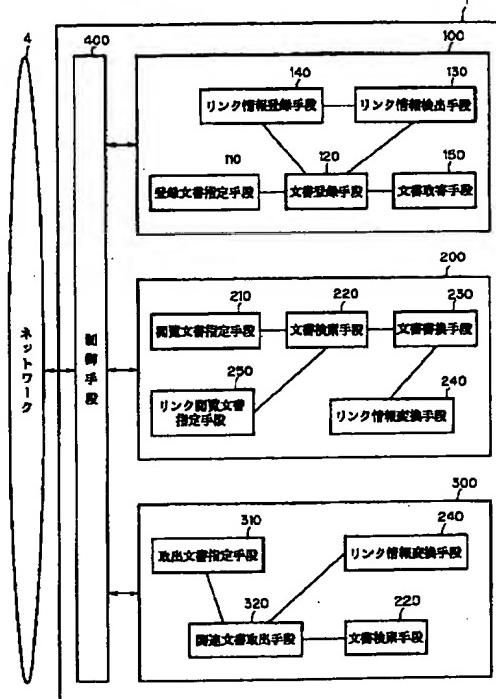
(71)出願人 000006747
株式会社リコー
東京都大田区中馬込1丁目3番6号
(72)発明者 今郷 詔
東京都大田区中馬込1丁目3番6号 株式
会社リコー内
(74)代理人 100079843
弁理士 高野 明近
Fターム(参考) 5B009 SA03 SA12 VA09
5B082 EA00 EA01

(54)【発明の名称】 文書登録システム、文書登録方法、その方法を実行させるための記録媒体、文書閲覧システム、文書閲覧方法、その方法を実行させるための記録媒体および文書取出しシステム

(57)【要約】

【課題】 文書データベースへ登録する文書情報を変更せずに、データベースへ登録した後でもそのままリンクを辿ることができるハイパー文書を登録・閲覧・取出しができるシステムを提供することを目的とする。

【解決手段】 文書登録システムは、登録すべき文書を指定する登録文書指定手段(110)と、前記登録文書指定手段(110)で指定された文書を解析しリンク情報を検出するリンク情報検出手段(130)と、文書データベースへ前記リンク情報検出手段(130)で対象となった文書を登録する文書登録手段(120)と、前記リンク情報検出手段(130)で検出したリンク情報を前記文書登録手段(120)で登録した文書と関連付けて管理するリンク情報登録手段(140)とを備える。



【特許請求の範囲】

【請求項1】 登録すべき文書を指定する登録文書指定手段と、前記登録文書指定手段で指定された文書を解析しリンク情報を検出するリンク情報検出手段と、文書データベースへ前記リンク情報検出手段で対象となった文書を登録する文書登録手段と、前記リンク情報検出手段で検出したリンク情報と前記文書登録手段で登録した文書とを関連付けて管理するリンク情報登録手段とを備えたことを特徴とする文書登録システム。

【請求項2】 請求項1記載の文書登録システムにおいて、前記リンク情報検出手段で検出したリンク情報をもとにリンク先の文書を取寄せる文書取寄手段を備え、前記リンク情報検出手段は、前記文書取寄手段で取寄せた文書も対象としてリンク情報を検出し、前記文書登録手段は、前記文書取寄手段で取寄せた文書も登録し、前記リンク情報登録手段は、前記文書取寄手段で取寄せた文書も対象として関連付けて登録するようにしたことを特徴とする文書登録システム。

【請求項3】 請求項1記載の文書登録システムにおいて、前記リンク情報検出手段は、指定されたリンク情報の表現パターンについてはリンク情報の検出対象から除外することを特徴とする文書登録システム。

【請求項4】 請求項1記載の文書登録システムにおいて、前記リンク情報検出手段は、指定されたリンク情報の表現パターンについてのみリンク情報の検出対象とすることを特徴とする文書登録システム。

【請求項5】 閲覧したい文書を指定する閲覧文書指定手段と、前記閲覧文書指定手段で指定された文書を検索して取出す文書検索手段と、リンク情報を文書データベースの検索情報へ変換するリンク情報変換手段と、前記文書検索手段で取出した文書に存在するリンク情報を前記リンク情報変換手段によって得られる対応した文書の検索情報へ書換えて文書を作成する文書書換手段と、前記文書書換手段で書換えた文書を閲覧中に該文書中のリンク情報を参照した場合、該リンク情報に対応する文書を取出すように指示するリンク閲覧文書指定手段とを備えたことを特徴とする文書閲覧システム。

【請求項6】 請求項5記載の文書閲覧システムにおいて、前記文書書換手段は書換えた文書をキャッシュしておき、次回からの閲覧時にはキャッシュから取り出すようにしたことを特徴とする文書閲覧システム。

【請求項7】 請求項6記載の文書閲覧システムにおいて、前記文書書換手段でのキャッシュは一定数又は一定期間保持するようにしたことを特徴とする文書閲覧システム。

【請求項8】 取出したい文書を指定する取出文書指定手段と、文書中にリンク情報がある場合、該リンク情報を文書データベースの検索情報へ変換するリンク情報変換手段と、前記リンク情報変換手段で得た検索情報によって文書を検索して取出す文書検索手段と、前記取出文

書指定手段で指定された文書を前記文書検索手段で取出し、その内容にリンク情報が存在する場合、該リンク情報を前記リンク情報変換手段で得た検索情報によって該リンク情報に対応した文書を取出す関連文書取出手段とを備えたことを特徴とする文書取出しシステム。

【請求項9】 登録すべき文書を指定し、この文書と該文書中のリンク先の文書を解析しリンク情報を検出し、検出したリンク情報と対応する文書とを関連付けて管理することを特徴とする文書登録方法。

【請求項10】 登録すべき文書を指定し、この文書と該文書中のリンク先の文書を解析しリンク情報を検出し、検出したリンク情報と対応する文書とを関連付けて管理するプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項11】 閲覧したい文書を指定し、指定された文書を検索して取出し、この文書に存在するリンク情報を文書データベースの検索情報へ書換えて文書を作成し、この書換えた文書を閲覧中に該文書中のリンク情報を参照した場合、該リンク情報に対応する文書を取出すように指示してリンク先の閲覧文書を指定するようにしたことを特徴とする文書閲覧方法。

【請求項12】 閲覧したい文書を指定し、指定された文書を検索して取出し、この文書に存在するリンク情報を文書データベースの検索情報へ書換えて文書を作成し、この書換えた文書を閲覧中に該文書中のリンク情報を参照した場合、該リンク情報に対応する文書を取出すように指示してリンク先の閲覧文書を指定するようにしたプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】 本発明は、文書データベースへの登録、閲覧および取出しシステムに関し、特にリンク情報をもつ文書の管理をするための技術に関する。

【0002】

【従来の技術】 現在、インターネット等のコンピュータネットワークにおいて流通する文書はHTML、SGML、XML等のハイパーテキストを使用して作成されている。このようなハイパーテキストによって作成された文書では、論理的には一つの文書が物理的には複数のファイルで構成される文書要素(ノード)に分割されている。このように複数の文書要素間の関係付けを行うことをリンクと言い、リンクによって複数の文書要素を統合した文書をハイパー文書という。あるノードから他のノードへの関係付け(リンク)を表現するには、参照先ノードのファイル名などを、そのノードが従う文書フォーマット固有の方法で指定する。例えば、HTMLでは、特定のタグの中のHREF属性等の値としてファイル名等(図13参照)を指定する。図13でのハイパー文書のリンクは、hub.html, node.html, /i

`image/img1.gif` の 3 つのファイルから構成されており、`hub.html` から `node.html` と `image/img1.gif` という 2 つのファイルへなされており、更に、`node.html` から `hub.html` へリンクされている。

【0003】一方、文書データベースなどを使った文書管理システムでは、文書を元のファイル名だけで特定することはできない。複数の異なる文書が、元は同じファイル名だったということもありうるためである。通常は、一つの文書には一つの一意な文書識別子（ID）が与えられ、その文書 ID を指定すれば対応する文書が一つだけ定まるように管理されている。例えば、図 14 は、図 13 のハイパー文書を文書データベースへ登録したときの様子を示しており、元の `hub.html` というファイルは `doc0001` という文書 ID で管理されることになる。ところが、登録されたノードの文書 (`doc0001`) の中では、元のファイル名 `node.html` が記されたままである。

【0004】

【発明が解決しようとする課題】一般には、このようなハイパー文書を文書データベースへ格納するとき、リンク情報は

- (1) そのまま変更せずに格納されるか（図 14 参照）
- (2) リンク情報を書き換えてから格納する（図 15 参照）

という 2 つの方法がとられている。

【0005】(1) の方法では、元のファイル名で文書データベースから文書を取出すことができないので、リンクが付いている文書のリンクを辿って別のノードをアクセスすることができなくなるという問題点がある。この方法である文書を閲覧するときには、これに関連した文書を文書データベースの中からすべて取出してから閲覧することになって 2 重手間となってしまうので効率が良くないし、ハイパー文書の構成要素を単独で利用することもできない。

【0006】又、(2) の方法では、相互にリンクされた一連の文書すべての登録が確定しないと文書 ID が確定しないため、2 パス処理となり、効率が悪くなるし、また、元文書にデジタル署名が付随している場合、文書内容を変更するとその署名が無効となってしまうという問題点がある。

【0007】このような問題点を解決するために本発明では、

- ・文書データベースへ登録する文書情報を変更せずに、
- ・データベースへ登録した後でもそのままリンクを辿ることができる

ハイパー文書を登録・閲覧・取出しができるシステムを提供することを目的とする。

【0008】

【課題を解決するための手段】請求項 1 の文書登録シ

テムは、登録すべき文書を指定する登録文書指定手段と、前記登録文書指定手段で指定された文書を解析しリンク情報を検出するリンク情報検出手段と、文書データベースへ前記リンク情報検出手段で対象となった文書を登録する文書登録手段と、前記リンク情報検出手段で検出したリンク情報と前記文書登録手段で登録した文書とを関連付けて管理するリンク情報登録手段とを備えた。

【0009】請求項 2 の文書登録システムは、請求項 1 記載の文書登録システムにおいて、前記リンク情報検出手段で検出したリンク情報をもとにリンク先の文書を取寄せる文書取寄手段を備え、前記リンク情報検出手段は、前記文書取寄手段で取寄せた文書も対象としてリンク情報を検出し、前記文書登録手段は、前記文書取寄手段で取寄せた文書も登録し、前記リンク情報登録手段は、前記文書取寄手段で取寄せた文書も対象として関連付けて登録するようにした。

【0010】請求項 3 の文書登録システムは、請求項 1 記載の文書登録システムにおいて、前記リンク情報検出手段は、指定されたリンク情報の表現パターンについてはリンク情報の検出対象から除外するようにした。

【0011】請求項 4 の文書登録システムは、請求項 1 記載の文書登録システムにおいて、前記リンク情報検出手段は、指定されたリンク情報の表現パターンについてのみリンク情報の検出対象とするようにした。

【0012】請求項 5 の文書閲覧システムは、閲覧したい文書を指定する閲覧文書指定手段と、前記閲覧文書指定手段で指定された文書を検索して取出す文書検索手段と、リンク情報を文書データベースの検索情報へ変換するリンク情報変換手段と、前記文書検索手段で取出した文書に存在するリンク情報を前記リンク情報変換手段によって得られる対応した文書の検索情報へ書換えて文書を作成する文書書換手段と、前記文書書換手段で書換えた文書を閲覧中に該文書中のリンク情報を参照した場合、該リンク情報に対応する文書を取出すように指示するリンク閲覧文書指定手段とを備えた。

【0013】請求項 6 の文書閲覧システムは、請求項 5 記載の文書閲覧システムにおいて、前記文書書換手段は書換えた文書をキャッシュしておき、次回からの閲覧時にはキャッシュから取り出すようにした。

【0014】請求項 7 の文書閲覧システムは、請求項 6 記載の文書閲覧システムにおいて、前記文書書換手段でのキャッシュは一定数又は一定期間保持するようにした。

【0015】請求項 8 の文書取出しシステムは、取出したい文書を指定する取出文書指定手段と、文書中にリンク情報がある場合、該リンク情報を文書データベースの検索情報へ変換するリンク情報変換手段と、前記リンク情報変換手段で得た検索情報によって文書を検索して取出す文書検索手段と、前記取出文書指定手段で指定された文書を前記文書検索手段で取出し、その内容にリンク

情報が存在する場合、該リンク情報を前記リンク情報変換手段で得た検索情報によって該リンク情報に対応した文書を取出す関連文書取出手段とを備えた。

【0016】請求項9の文書登録方法は、登録すべき文書を指定し、この文書と該文書中のリンク先の文書を解析しリンク情報を検出し、検出したリンク情報と対応する文書とを関連付けて管理することを特徴とする文書登録方法。

【0017】請求項10のコンピュータ読み取り可能な記録媒体は、登録すべき文書を指定し、この文書と該文書中のリンク先の文書を解析しリンク情報を検出し、検出したリンク情報と対応する文書とを関連付けて管理するプログラムを記録した。

【0018】請求項11の文書閲覧方法は、閲覧したい文書を指定し、指定された文書を検索して取出し、この文書に存在するリンク情報を文書データベースの検索情報へ書換えて文書を作成し、この書換えた文書を閲覧中に該文書中のリンク情報を参照した場合、該リンク情報に対応する文書を取出すように指示してリンク先の閲覧文書を指定する。

【0019】請求項12のコンピュータ読み取り可能な記録媒体は、閲覧したい文書を指定し、指定された文書を検索して取出し、この文書に存在するリンク情報を文書データベースの検索情報へ書換えて文書を作成し、この書換えた文書を閲覧中に該文書中のリンク情報を参照した場合、該リンク情報に対応する文書を取出すように指示してリンク先の閲覧文書を指定するようにしたプログラムを記録した。

【0020】

【発明の実施の形態】以下、4つの実施例を用いて、本発明の構成および動作原理について説明する。

【0021】1. 第1の実施例

本発明の第1の実施例について図1～8を用いて説明する。

(1) 全体の環境

図1は、文書登録システム、文書閲覧システムおよび文書取出しシステムの稼働環境を示す図である。リンクサーバー1は、本発明の文書登録システム、文書閲覧システムおよび文書取出しシステムが稼動するコンピュータである。文書サーバー2は、ユーザーからの登録・閲覧・取出し要求に応じるための文書（リンク付き文書も含む）を保持しておく文書データベースが稼動するコンピュータであり、リンクサーバー1からの要求に応じて動作する。クライアント3は、ネットワーク上のデータソース5や自分のコンピュータ上にあるリンク付きの文書を文書サーバー2の文書データベースへ登録・閲覧・取出しの指示をするユーザーのコンピュータである。クライアント3は、単独でも複数あっても良い。ネットワーク4は、リンクサーバー1、文書サーバー2、複数のクライアント3と複数のデータソース5とを結合するネット

ワークであり、LAN（Local Area Network）であっても、WAN（Wide Area Network）であっても、インターネットのようなものであっても良い。データソース5は、ユーザーが文書サーバー2の文書データベースへ登録したいリンク付きの文書が存在するネットワーク上のコンピュータである。データソース5は、クライアント3自身であっても良い。

【0022】図2は、本発明が稼動するリンクサーバー1、文書サーバー2およびクライアント3のコンピュータハードウェア構成を示す図である。入力手段10はキーボード、マウス、タッチパネル等により構成され、情報の入力に使用される。出力手段20は、種々の出力情報や入力手段10からの入力された情報などを出力させるものであって、表示装置やプリンタ等から構成される。CPU30は、種々のプログラムを動作させる。メモリ40は、プログラムがCPU30によって実行されるときに一時的に作成される情報やプログラム等を保持する。記憶手段50は、文書データベース、プログラムやプログラムの実行時の一時的な情報等を保持する。媒体駆動装置60は、プログラムやデータ等を記憶した記録媒体を装着してそれらを読み込み、メモリ40または記憶手段50へ格納するのに用いられる。ネットワークインターフェース（I/F）70は、コンピュータとネットワーク4とのインターフェースをとつて、ネットワーク上の情報の送受信を行うためのインターフェース装置である。

【0023】(2) リンク情報を管理する表のデータ構造

図3は、ユーザーが登録したいリンク付きの文書に含まれているリンク情報を文書サーバー2の文書データベースへ登録されたときの文書識別子（ここでは文書IDを識別子とする）にどのように対応するかを管理するリンク情報対応表80のデータ構造を表現したものである。このリンク情報対応表80は、文書名リスト81および1文書に対応したリンク情報リスト82からなっている。文書名リスト81は、リンク付きの文書とその文書からリンクされた文書を1つにまとめて文書名を付けて文書データベースに登録し、その各まとまりである文書名に何があるかを列挙しており、次の2つの項目がある。

- ・まとまりとしての文書名
- ・この文書に対するリンク情報リスト82へのポインタまた、リンク情報リスト82は1つの文書ごとに次の項目がある。
 - ・文書データベースに登録したときの文書識別子（文書ID）
 - ・この文書のものとのファイル名（リンク位置があるときには未定義）
 - ・リンク位置（これは次の3つの情報からなる）

- ・リンク情報があるファイルの先頭からの文字位置
- ・リンク情報の文字数
- ・リンク情報の文字列
- ・対応文書の識別子（対応文書の文書ID）

【0024】例えば、図3を例にとると、文書名リスト81には次のものが設定される。

- ・「ハブ文書」という文書名
- ・この「ハブ文書」に対するリンク情報リスト82へのポインタ

【0025】リンク情報リスト82の各項目は、次のように設定される。文書データベースに登録された文書自身に関しては、上記の項目のうち最初の2つについてのみ登録される。例えば、図13のhub.htmlというファイルが識別子doc0001として文書データベースに登録されたとすると、この文書の項目は次のようにになる。

- ・文書データベースに登録したときの文書識別子→doc0001
- ・この文書のもとのファイル名→hub.html
- ・リンク位置→未定義
- ・対応文書の識別子→未定義

【0026】また、文書の中にリンク情報があるときはそれぞれについてすべての情報が設定される。例えば、図8のhub.htmlというファイルにある2つのリンク情報は、それぞれ次のように設定される。ファイルnode.htmlが識別子doc0002、ファイル/image/img1.gifが識別子doc0003として文書データベースへ登録されたとする。

- (a) ファイルnode.htmlへのリンク情報
- ・文書データベースに登録したときの文書識別子→doc0001
 - ・この文書のもとのファイル名→未定義
 - ・リンク位置（これは次の3つの情報からなる）
 - ・リンク情報があるファイルの先頭からの文字位置→49
 - ・リンク情報の文字数→9
 - ・リンク情報の文字列→node.html
 - ・対応文書の識別子→doc0002

- (b) ファイル/image/img1.gifへのリンク情報
- ・文書データベースに登録したときの文書識別子→doc0001
 - ・この文書のもとのファイル名→未定義
 - ・リンク位置（これは次の3つの情報からなる）
 - ・リンク情報があるファイルの先頭からの文字位置→19
 - ・リンク情報の文字数→15
 - ・リンク情報の文字列→/image/img1.gif
 - ・対応文書の識別子→doc0003

【0027】(3) 本発明を含めた文書の登録から閲覧・取出しの全体の流れ

本発明を説明する前に、クライアントからの要求がどのように処理されていくかを図4を用いて概説する。

(A) ユーザーは、文書管理システム（図示されていない）をクライアント3で起動する。文書管理システムは、文書サーバー2の文書データベースに登録されている文書の一覧をリンクサーバー1から得て出力手段20へ出力する。ユーザーは出力結果からこれから登録するリンク付きの文書が登録されていないことを確かめてから、入力手段10から登録のボタンを押す。

(B) クライアント3の記憶手段50に格納されているファイルを表示させてその中から登録したいファイルを選択する。（ここでは、hub.htmlを選択した。）

(C) 文書サーバー2の文書データベースへ登録する文書名（「ハブ文書」とする）と書誌事項等を入力する。この入力によってクライアント3は、リンクサーバー1へここで指定したファイルを指定した文書名で登録作業を行う。この登録作業が本発明の文書登録システムである。

(D) この登録作業が終了すると、リンクサーバー1は指定通りに文書サーバー2へ登録したことをクライアント3の文書管理システムへ戻す。文書管理システムは、登録された文書名を出力手段20へ出力する。

【0028】又、このようにして登録された文書を閲覧するには、(A)と同様にして文書管理システムを起動して、リンクサーバー1（文書サーバー2の文書データベース）に登録された文書の一覧を表示させる。この中から例え今登録した「ハブ文書」を選択する。

(E) 選択された「ハブ文書」の中の構成要素が表示される。この中のファイルで自分のコンピュータ上へコピーリーしたいときには閲覧ボタン又は保存ボタンを押す。

(F) 閲覧ボタンを押したときには、指定された構成要素のファイル名をリンクサーバー1へ送り、文書サーバー2の文書データベースから指定した文書を取出して、クライアント3へ送り返してくれる。その後、送り返された文書を閲覧中にリンク情報を参照すると、この参照されたリンク情報をリンクサーバー1へ送り、文書サーバー2の文書データベースから対応する文書を取出してクライアント3へ送り返してくれる。この閲覧作業が本発明の文書閲覧システムである。

【0029】(G) また、(E)で保存ボタンを押したときには、クライアント3の保存先のフォルダ名を入力すると、文書サーバー2の文書データベースから対応する文書をすべて取出してクライアント3へ送り返し、文書管理システムはそれを格納する。この取出し作業が本発明の文書取出しシステムである。

【0030】図5は、リンクサーバー1で稼動する各システム（文書登録システム100、文書閲覧システム200および文書取出しシステム300）およびその制御手段400からなる機能構成図である。図6は、リンク

サーバー1で稼動する3つのシステムを統合的に制御する制御手段400の処理の流れを説明するフローチャートである。

【0031】ステップS1：クライアント3のユーザーからの指示を受信する。

ステップS2：受信した指示が「ファイルの登録」であれば、ステップS3へ、そうでなければステップS4へ分岐する。

ステップS3：受信したユーザーの指定したファイルを登録するために「文書登録システム」を呼び出す。

この結果の状態をもって、ステップS11へ進む。この処理状態には、次のようなものがある。

- ・正常に終了した場合、登録された文書名とその構成要素の文書IDを戻す。

- ・異常終了した場合、その理由を戻す。

ステップS4：受信した指示が「ファイルの閲覧」であれば、ステップS5へ、そうでなければステップS6へ分岐する。

ステップS5：受信したユーザーの指定したファイルを閲覧するために「文書閲覧システム」を呼び出す。取出された閲覧文書をもって、ステップS11へ進む。

ステップS6：受信した指示が「リンク情報から示されたファイルの閲覧」であれば、ステップS7へ、そうでなければステップS8へ分岐する。

ステップS7：受信したユーザーの指定したリンクファイルを閲覧するために「文書閲覧システム」を呼び出す。取出された閲覧文書をもって、ステップS11へ進む。

ステップS8：受信した指示が「ファイルの取り出し」であれば、ステップS9へ、そうでなければステップS10へ分岐する。

ステップS9：受信したユーザーの指定したファイルに従属したファイル全部を取出すために「文書取り出しシステム」を呼び出す。取出された文書をもって、ステップS11へ進む。

ステップS10：ここは例えば、クライアント3のユーザーから指定されたリンクサーバー1内で管理されているすべての文書名とその文書を構成するリンク文書およびそれらの文書IDとをリンク情報対応表80から抽出する。これをもって、ステップS11へ進む。

ステップS11：クライアント3のユーザーへ要求された処理結果を送信し、次の受信を待つ。

【0032】これら3つのシステムについて、以下に詳細な説明をする。

(4) 文書登録システム100

図5のリンクサーバー1の機能構成図が示すように文書登録システム100は、登録文書指示手段110、文書登録手段120、リンク情報検出手段130、リンク情報登録手段140、文書取寄手段150とから構成される。登録文書指示手段110は、クライアント3のユー

ザーが指定した登録する文書名と登録する文書の内容を受け取る。

【0033】文書登録手段120は、登録文書指示手段110が受け取った登録文書の内容を文書サーバー2へ送信して文書データベースへ登録してもらい、その文書IDを受け取り、リンク情報登録手段140へ送る。リンク情報検出手段130は、指定された文書の内容にリンク情報があるかどうかを調べる。検出された場合、もとになった文書と検出されたリンク情報を対応させてリンク情報対応表80へ登録するようにリンク情報登録手段140へ依頼する。また、リンク情報に記述された文書を文書取寄手段150に依頼して取寄せる。取寄せられた文書に対して、文書登録手段120で文書サーバー2へ登録するとともに、リンク情報を検出しリンク情報対応表80へ登録するようにリンク情報登録手段140へ依頼する。なお、リンク情報検出手段130は、検出すべきでないリンク情報の表現パターンを用意しておき、そのパターンにマッチするリンクは検出しないようすることもできる。例えば、"http://x/y/*"というパターンにマッチ（"*"は任意の文字列）するリンクは検出しない。逆に、検出すべきリンク情報の表現パターンを用意しておき、それにマッチしたリンクだけを検出するように構成することもできる。例えば、"file:///*"というパターンにマッチするリンクだけを検出する。

【0034】リンク情報登録手段140は、文書登録された文書をリンク情報対応表80へ登録する。またリンク情報検出手段130から検出されたリンク情報も呼び出しもとの文書とリンク先の文書とを関連付けてリンク情報対応表80へ登録する。文書取寄手段150は、リンク情報検出手段130で検出された新たな文書に対してクライアント3のコンピュータ又はネットワーク4上のデータソース5から取寄せる。この取寄せられた文書はまた、リンク情報検出手段130によってリンク情報を検出される。

【0035】文書登録システムの処理の流れを図7のフローチャートを用いて説明する。

ステップS21：ユーザーから指定されたファイルを登録する場合は、この指定されたファイルを文書サーバー2へ送信し、文書データベースへ登録してもらう。この登録したときの「文書ID」を文書サーバー2から受信する。ファイルの中にあったリンク情報によって取り寄せられたファイルについても同様に処理する。

【0036】ステップS22：この「文書ID」と「ファイルの名前」をリンク情報情報対応表80へ次のようにして登録する。まず、ユーザーによって指定された「文書名」によって文書名リスト81を探し、登録されていないときには「文書名」を登録し、リンク情報リスト82へのポインタを設定する。文書名リスト81に登録されれば、リンク情報リスト82へのポインタを取出す。このリンク情報リスト82へ、文書サーバー2

から得た「文書ID」と「そのファイルの名前」を登録する。例えば、図13のhub.htmlファイルの場合、図8の第1エントリーのように、文書IDをdoc0001、元のファイルの名前をhub.htmlとして登録される。更に、この登録したファイルがすでに処理されているリンク情報リスト82の中のリンク情報のうち対応文書IDが未定義となっているものを見つけて「文書ID」を設定する。図13のhub.htmlファイルの場合、図8の第3エントリーの対応文書IDは、第4エントリーの"node.html"が文書サーバー2へ登録されたとき文書ID(doc0002)を得てこれを設定する。

【0037】ステップS23：今登録したファイルの内容からリンク情報を検出する。リンク情報は、例えばHTML文書の場合、図13のように<a>タグ、タグ等にあるので、このパターンについて分析することによって得られる。図13のHub.htmlファイルの場合には、リンク情報として次の2つが得られる。

- ・ファイルの先頭より49文字目から9文字長の長さで、"node.html"という文字列と、
- ・ファイルの先頭より19文字目から15文字長の長さで、"/image/img1.gif"という文字列。

なお、検出すべきでないリンク情報の表現パターンを用意しておき、そのパターンにマッチするリンクは検出しないようにすることもできる。例えば、"http://x/y/*"というパターンにマッチ("*"は任意の文字列)するリンクは検出しない。逆に、検出すべきリンク情報の表現パターンを用意しておき、それにマッチしたリンクだけを検出するように構成することもできる。例えば、"file:///*"というパターンにマッチするリンクだけを検出する。

【0038】ステップS24：上記のステップで検出されたものをすべてリンク情報リスト82へ登録する。登録は、リンク情報が記述されているファイルの「文書ID」と検出したリンク情報を登録する。図13のhub.htmlの場合は、図8の第2、第3エントリーのように登録する。一般に対応文書IDは、この時点では未定義であるが、すでに文書サーバー2へ登録されれば、リンク情報リスト82に定義されているので、それを設定する。

【0039】ステップS25：上記のステップで登録されたリンク情報のそれぞれがサーバー2へ登録されているかどうかをチェックする。すべて登録されれば終了する。登録されていないものがあれば、ステップS26へ進む。

ステップS26：まだリンク情報リスト82のリンク情報に対して文書サーバー2に登録されていないリンク文書名をリンク位置の文字列から取出して、そのファイルの内容を取寄せる。例えば、図8の第3エントリーのように"node.html"となっているときにはユーザーのクラ

イアント3のコンピュータから対応するファイルの内容を送ってもらう。又、"http://www.xyz.co.jp/node.htm1"のようになっている場合は、ネットワーク4上のデータソース5から取寄せる。

ステップS27：今取寄せたファイルに対して、再帰的にこの「文書登録」の処理を実行する。この再帰的な処理から戻ったときには、次の未処理のリンク情報の処理のためにステップS25へ戻る。

【0040】(5) 文書閲覧システム200

図5のリンクサーバー1の機能構成図が示すように文書閲覧システム200は、閲覧文書指示手段210、文書検索手段220、文書書換手段230、リンク情報変換手段240、リンク閲覧文書指示手段250とから構成される。閲覧文書指示手段210は、クライアント3のユーザーが指定した閲覧したいファイルの文書IDを受け取る。ここではクライアント3のユーザーインターフェースによってリンクサーバー1中に登録されている文書一覧の中から選択された文書を構成するファイルを選択したときに特定されたファイルの文書IDを指定して閲覧を申し込んでくるものと仮定した。閲覧したい文書名とその構成要素のファイル名とを指定するようにしても良い。このときはこれらの情報からリンク情報対応表80を検索することによって文書IDを得ることができ。文書検索手段220は、閲覧文書指示手段210が受け取った閲覧したいファイルの文書IDを文書サーバー2へ送信して文書データベースからファイルの内容を受け取る。閲覧文書指示手段210へ送る。

【0041】文書書換手段230は、受理したファイルの内容にリンク情報があるかどうかリンク情報対応表80を使って調べる。例えば、「file://foo.html」というようなリンク情報があると、クライアント3はリンクサーバー1を経由することなく、クライアント3上のファイル「foo.html」を探そうとするから、ファイルをクライアント3へ送信する前に、ファイルの内容のリンク情報を書き換える。例えばクライアント3をユーザーのコンピュータ上で動作するWebブラウザ、リンクサーバー1（リンクサーバーのホスト名を「linkserver」とする）の制御手段400にWWWサーバーを利用したとして、リンク閲覧文書指示手段250をcgi(common gateway interface)プログラム「getdoc」によって構成したとすると、図13のhub.htmlファイルのは、図8からファイル"/image/img1.gif"の文書IDはdoc0003、ファイル"node.html"の文書IDはdoc0002であるので次のように書換えられる。

```
<html>

See <a href="http://linkserver/getdoc?doc0002">the
node</a>for details.
</html>
```

このように書換えると、クライアント3上のWWWブラウザで、リンクサーバー1から返された文書を表示し、リンク個所をクリックすると、リンクサーバー1のCGIプログラムを呼び出すことになるので、文書サーバー2上の該当ファイルを得ることになる。

【0042】リンク情報変換手段240は、文書検索手段230で得たファイルの内容にあるリンク情報をリンク情報対応表80から見つけ、リンク先のファイルの代わりに対応文書IDを知らせる。これをもとに上記のようなリンク先の書換えを行う。リンク閲覧文書指示手段250は、文書書換手段230で書換えたリンクを選択したときに動作し、この書換えによって埋め込まれた文書IDをもとに文書検索手段220へ送る。

【0043】又、文書書換手段230は、書換え後の文書を一定数あるいは一定時間キャッシュしておき、キャッシュされている文書に対してはこの書換え処理を省略するように構成することもできる。

【0044】文書閲覧システムの処理の流れを図9のフローチャートを用いて説明する。

ステップS31：ユーザーがクライアント3から閲覧したい文書の中の構成要素のファイルの文書IDを指定して閲覧を指示てくる。この閲覧したいファイルの文書IDを文書サーバー2へ送信して文書データベースからファイルの内容を受け取る。

ステップS32：受理したファイルの内容にリンク情報があるかどうかリンク情報対応表80を使って調べる。リンク情報を持っている場合、リンク情報リスト82から分かるリンク情報の位置の文字列を「対応文書ID」によって上記の文書書換手段230で説明したように書換え、処理を終了する。

ステップS33：上記のように書換えた後、クライアント3上の表示されたリンク情報を参照するとこの書換え情報によりリンク文書閲覧の指示がリンクサーバー1へなされる。書換え情報にもとづいて文書サーバー2からファイルを得るためにステップS31へ進む。

【0045】(6) 文書取り出しシステム300

図5のリンクサーバー1の機能構成図が示すように文書取り出しシステム300は、取出文書指示手段310、関連文書取出手段320、文書検索手段220、リンク情報変換手段240とから構成される。取出文書指示手段310は、クライアント3のユーザーが指定した取出したいファイルの文書IDを受け取る。ここではクライアント3のユーザーインターフェースによってリンクサーバー1中に登録されている文書一覧の中から選択された文書を構成するファイルを選択したときに特定されたファイルの文書IDを指定して取出しを申し込んでくるものと仮定した。取出したい文書名とその構成要素のファイル名とを指定するようにしても良い。このときはこれらの情報からリンク情報対応表80を検索することによって文書IDを得ることができる。

【0046】関連文書取出手段320は、取出文書指示手段310が受け取った取出したいファイルの文書IDを文書サーバー2へ送信して文書データベースからファイルの内容を受け取る。この受け取ったファイルをクライアント3へ送信したあと次のことを行う。この受け取ったファイルにリンク情報があるかをリンク情報対応表80によって探し、そのリンク情報にある対応文書IDをリンク情報変換手段240によって取出す。次に、この取出した対応文書IDのファイルを文書検索手段220を使って取出し、クライアント3へ送信する。この2つのステップをリンク情報対応表80に登録されている関連文書が無くなるまで繰り返す。

【0047】文書取り出しシステムの処理の流れを図10のフローチャートを用いて説明する。

ステップS41：クライアント3のユーザーが指定した取出したいファイルの文書IDを受け取り、ファイルの文書IDを文書サーバー2へ送信して文書データベースからファイルの内容を受け取る。

ステップS42：この受け取ったファイルをそのまま（書き換えもせずに）クライアント3へ送信する。

ステップS43：この受け取ったファイルに未処理のリンク情報があるかをリンク情報対応表80によって探し、なければ終了する。あればステップS44へ進む。

ステップS44：この受け取ったファイルのリンク情報にある対応文書IDを1つリンク情報リスト82から取出し、再帰的にこの「関連文書取り出し」の処理を実行する。この再帰的な処理から戻ったときには、次の未処理のリンク情報の処理のためにステップS43へ戻る。

【0048】2. その他の実施例

本発明は上記のように各サーバーを独立のコンピュータとして実現したものに限定されたものではなく、図11や図12に示したようにサーバーを統合して扱ってもいい。図11は、本発明の稼働環境をリンクサーバー1と文書サーバー2とを同一のコンピューターで実現したものである。又、図12は、本発明の稼働環境をリンクサーバー1、文書サーバー2およびクライアント3とを同一のコンピューターで実現したものである。

【0049】3. プログラムとしての実施例

更に、本発明は上記の実施形態のみに限定されたものではない。例えば、図5に示した各システム（文書登録システム、文書閲覧システム、文書取り出しシステム）を構成する各手段の各機能をプログラム化し、予めCD-ROM等の記録媒体に書き込んでおき、このCD-ROMをCD-ROMドライブを搭載したコンピュータに装着して、コンピュータへそのプログラムをロードすることによって、実施例の実施形態と同様な機能を実現することができる。ここで各システムを1つの統合されたプログラムとしても良いし、又はそれぞれ個々の単独のプログラムとしても良い。

【0050】尚、記録媒体としては半導体媒体（例え

ば、ROM、ICメモリカード等)、光媒体(例えば、DVD-ROM、MO、MD、CD-R等)、磁気媒体(例えば、磁気テープ、フレキシブルディスク等)のいずれであってもよい。

【0051】また、本発明の機能を実現するプログラムは、媒体の形で提供されるのみならず、通信によって提供されるものであっても良い。

【0052】

【発明の効果】以上説明したように特許請求の範囲の発明によれば、リンク情報が変更せずに文書データベースに保持されるので、文書データベース内のリンク付き文書のリンクも辿ることが可能になる。

【0053】又、文書データベース内の元の文書内容自体は変わらないので、デジタル署名が付いていても、その署名の有効性を正しく検証することができる。

【0054】又、文書データベース内の文書を閲覧時に閲覧結果をキャッシュするので、アクセスが頻繁な文書の処理速度が向上し、ユーザに待ち時間を与えない。

【0055】又、リンク付きの文書の構成文書のなかで、他のコンピュータシステムで管理されていたり、更新が激しい場合などで、登録すべきでない文書へのリンク情報のパターンを指定できるので、常に適切な構成文書だけを文書データベースとして管理することができる。

【図面の簡単な説明】

【図1】 本発明の文書登録システム、文書閲覧システムおよび文書取出しシステムの稼働環境を示す図である。

【図2】 本発明が稼働するコンピュータ環境を示すコンピュータハードウェア構成を示す図である。

【図3】 登録したいリンク付きの文書とリンク情報の対応を管理するリンク情報対応表のデータ構造を示す図である。

【図4】 文書の登録から閲覧・取出しの要求がどのように処理されていくかを説明するための画面例である。

【図5】 リンクサーバーで稼働する本発明の文書登録システム、文書閲覧システムおよび文書取出しシステム

およびその制御手段からなる機能構成図である。

【図6】 リンクサーバーで稼働する3つのシステムを統合的に制御する制御手段の処理の流れを説明するフローチャートである。

【図7】 本発明の文書登録システムの処理の流れを説明するフローチャートである。

【図8】 リンク情報リストの格納例を説明するための図である。

【図9】 本発明の文書閲覧システムの処理の流れを説明するフローチャートである。

【図10】 本発明の文書取出しシステムの処理の流れを説明するフローチャートである。

【図11】 本発明の第2のシステム環境を示す図である。

【図12】 本発明の第3のシステム環境を示す図である。

【図13】 リンク付きの文書の例である。

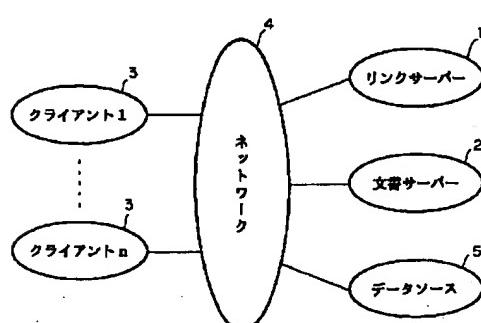
【図14】 リンク付きの文書を文書データベースへそのまま格納した例を示す。

【図15】 リンク付きの文書の望ましい登録例を示す図である。

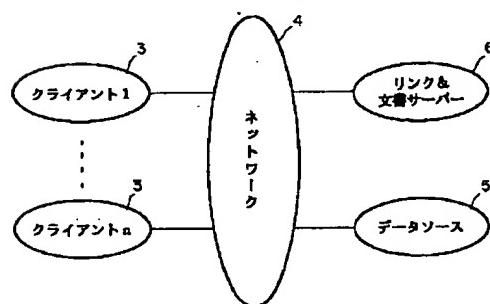
【符号の説明】

1…リンクサーバー、2…文書サーバー、3…クライアント、4…ネットワーク、5…データソース、10…入力手段、20…出力手段、30…CPU、40…メモリ、50…記憶手段、60…媒体駆動装置、70…インターフェース、80…リンク情報対応表、81…文書名リスト、82…リンク情報リスト、100…文書登録システム、110…登録文書指定手段、120…文書登録手段、130…リンク情報検出手段、140…リンク情報登録手段、150…文書取寄手段、200…文書閲覧システム、210…閲覧文書指定手段、220…文書検索手段、230…文書換手手段、240…リンク情報変換手段、250…リンク閲覧文書指定手段、300…文書取出しシステム、310…取出文書指定手段、320…関連文書取出手段、400…制御手段。

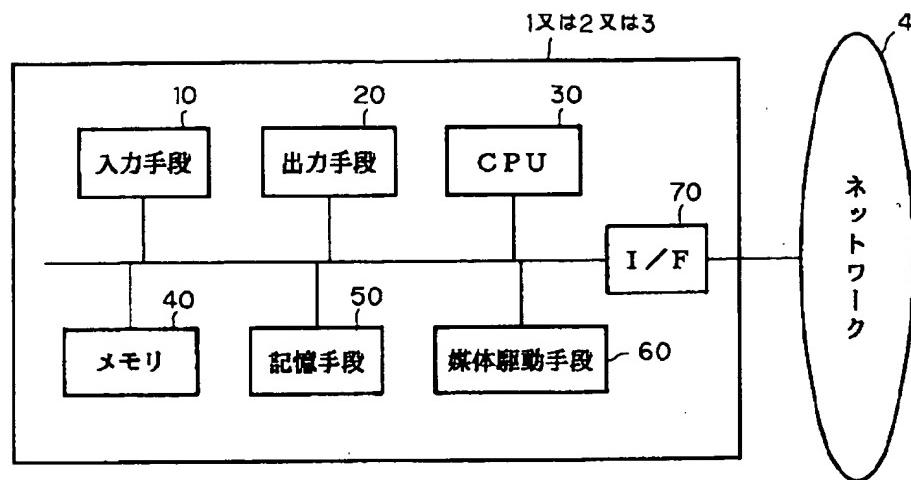
【図1】



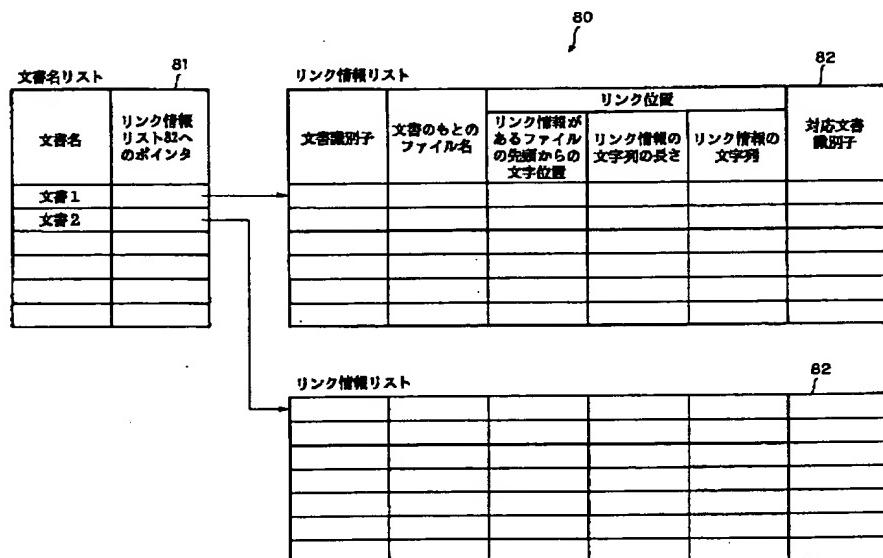
【図11】



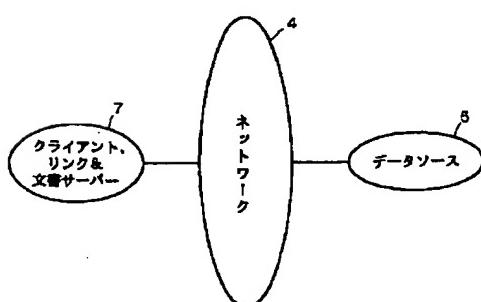
【図2】



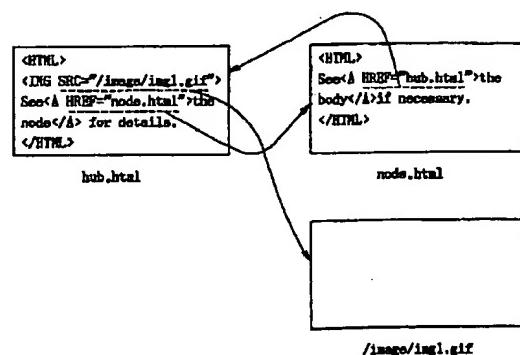
【図3】



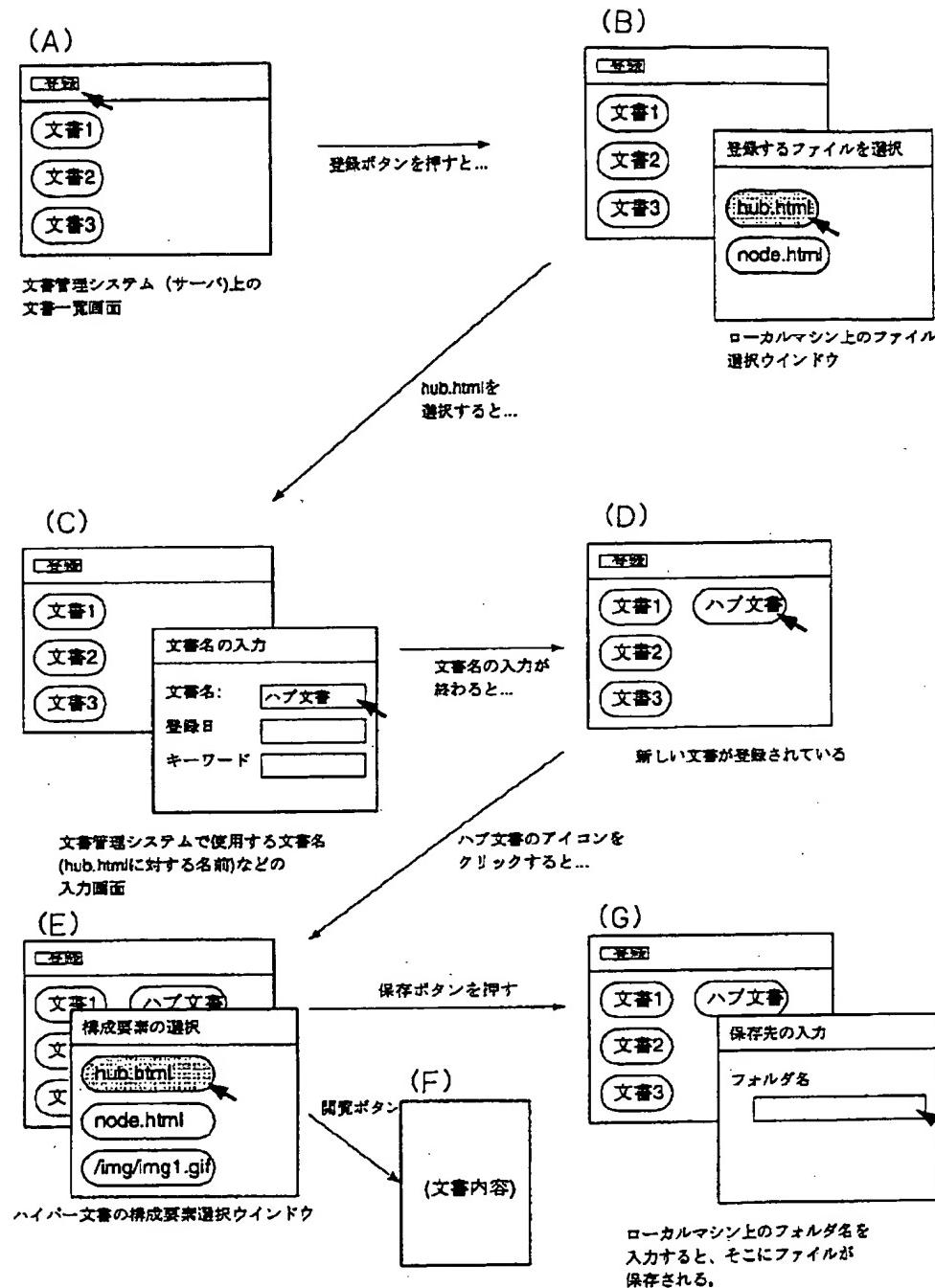
【図12】



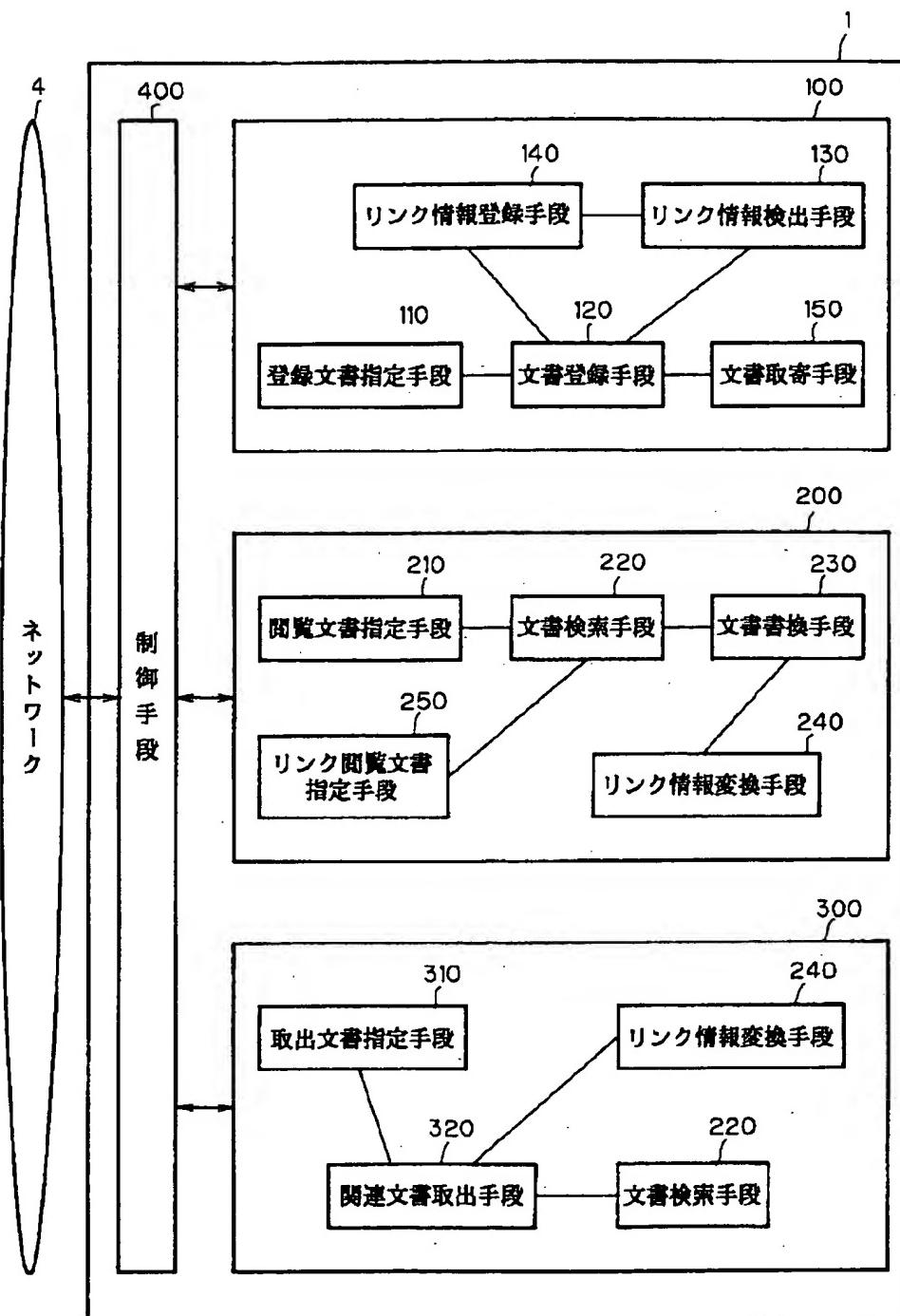
【図13】



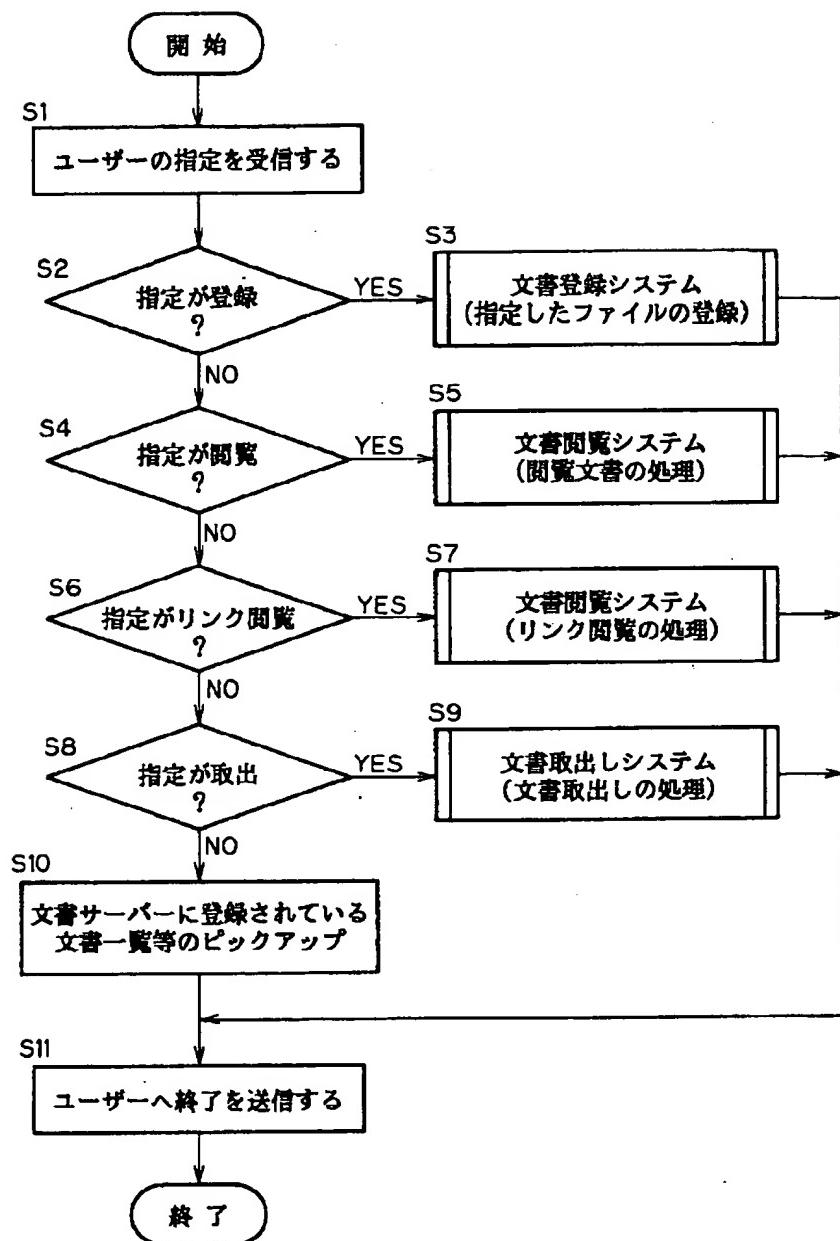
【図4】



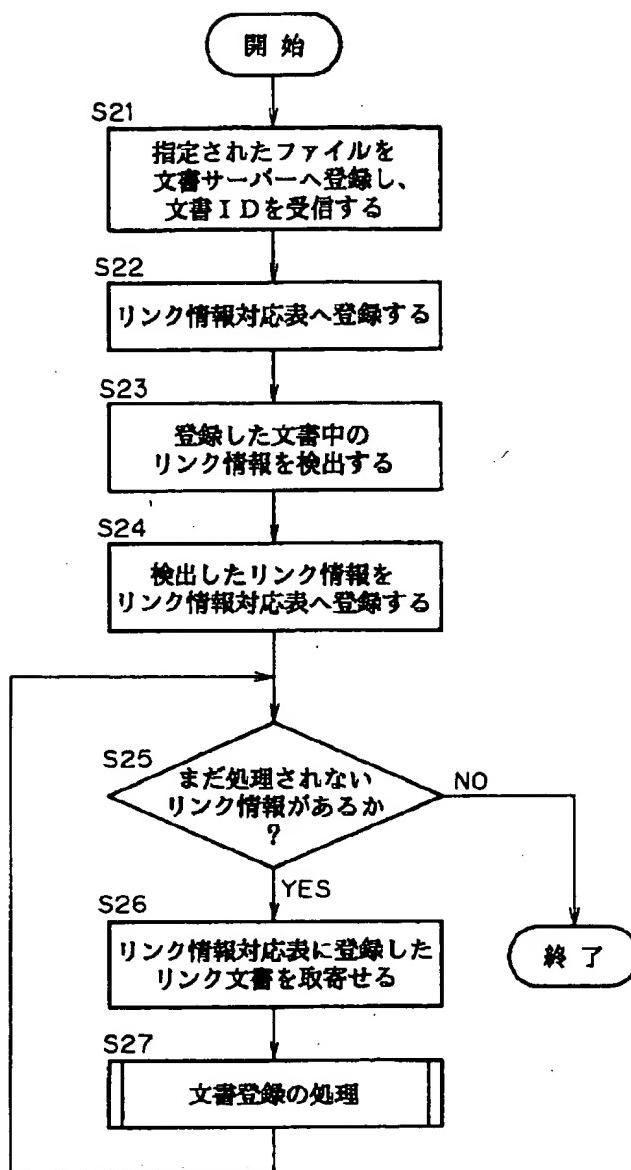
【図5】



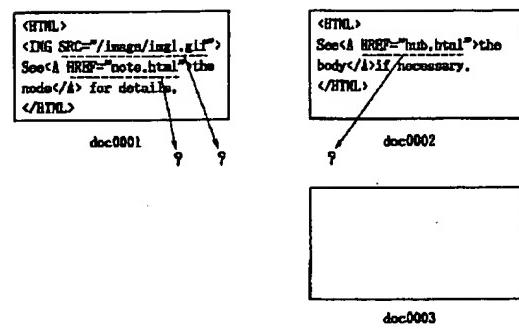
【図6】



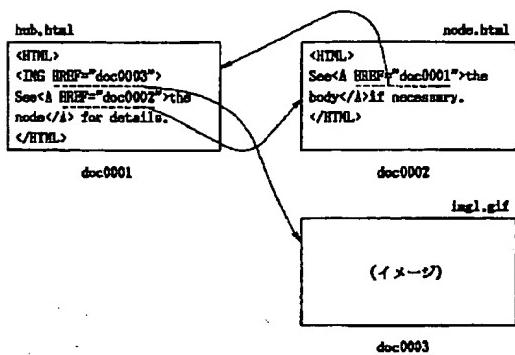
【図 7】



【図 14】



【図 15】

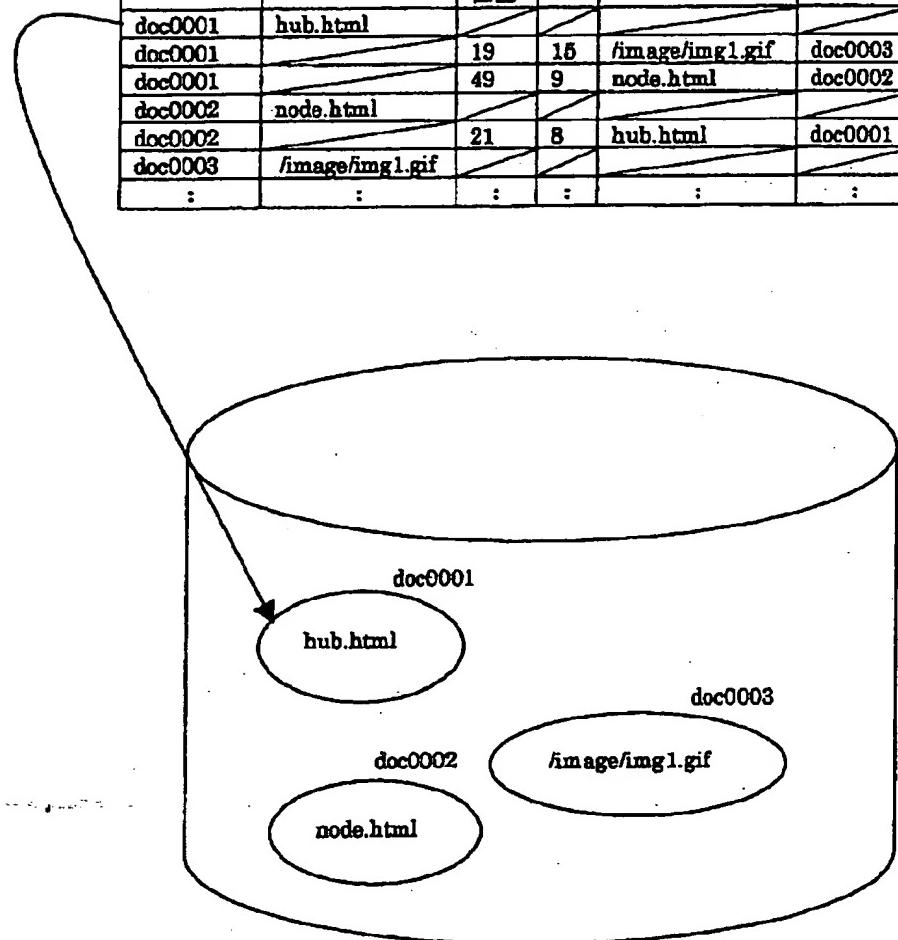


【図8】

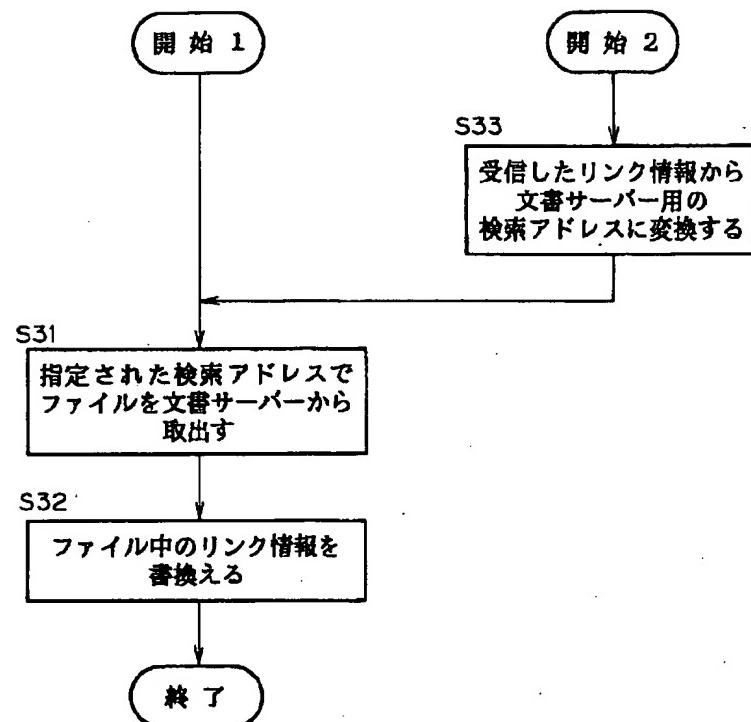
82

リンク情報リスト

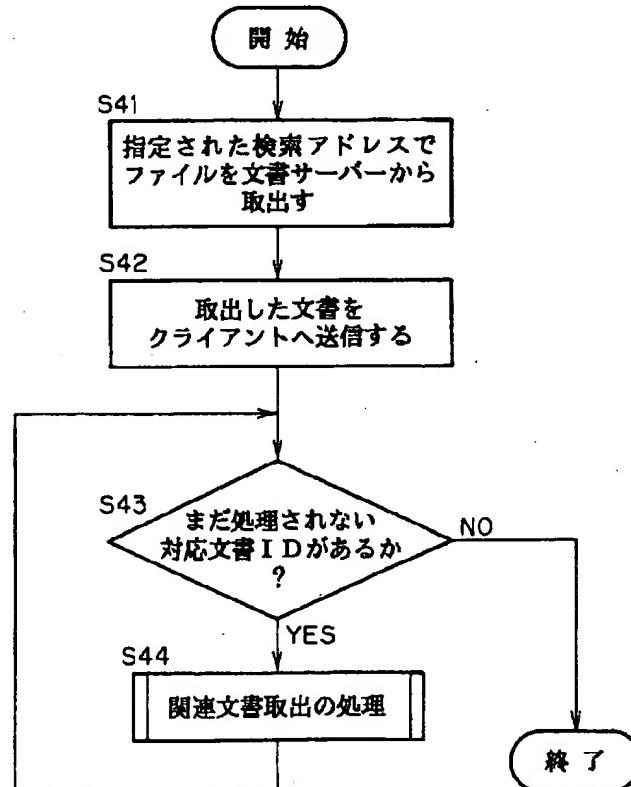
文書ID	文書のもとのファイル名前	リンク情報位置			対応文書ID
		ファイルの先頭から的位置	長さ	文字列	
doc0001	hub.html				
doc0001		19	15	/image/img1.gif	doc0003
doc0001		49	9	node.html	doc0002
doc0002	node.html				
doc0002		21	8	hub.html	doc0001
doc0003	/image/img1.gif				
:	:	:	:	:	:



【図9】



【図10】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.